

相対効率にもとづく リッジ係数の選択アルゴリズム

Algorithms for Selecting The Ridge Coefficient Based on Relative Efficiency

地道正行*

We consider some algorithms for selecting the ridge coefficient based on the relative efficiency of the ordinary ridge regression estimator to the ordinary least squares estimator. We give some Monte Carlo simulations for them with two kind of data sets by Hoerl (1962), and evaluate the performance of these algorithms by comparing some empirical criteria of them.

Masayuki Jimichi

JEL : C13, C15, C21, C44

キーワード：多重共線性、通常最小自乗推定量、通常リッジ回帰推定量、リッジ係数、相対効率

Keywords : Multicollinearity, Ordinary Least Squares Estimator, Ordinary Ridge Regression Estimator, Ridge Coefficient, Relative Efficiency

1 はじめに

通常リッジ回帰推定量は、線形回帰モデルにおいて多重共線性が存在する場合に回帰係数の推定精度を改良するために Hoerl and Kennard (1970) によって提案されたものである。通常リッジ回帰推定量を実際に利用する際には、リッ

* Masayuki Jimichi (Ph.D.) is Professor of Statistics, School of Business Administration, Kwansei Gakuin University, Nishinomiya, Japan.

E-mail: jimichi@kwansei.ac.jp

ジ係数 (k で表される.) を選択する必要があるけれども、この選択に関するアルゴリズムは現在まで非常に多くのものが提案されており、それらの良さの比較に関する多くの研究がある。(たとえば, Gibbons (1981) 参照.) これらのアルゴリズムは、さまざまな視点から提案されており、たとえば総平均 2 乗誤差などのある種の危険関数 (risk function) にもとづくものは以下の 2 種類に大別される:

- (M1) まず未知パラメータを含んだ総平均 2 乗誤差を正確に (または近似的に) 最小にするリッジ係数 $k = k^*$ を選択しておいて、次にその k^* を推定する.
- (M2) まず総平均 2 乗誤差を推定し、次にそれを正確に (または近似的に) 最小にする $k = k^*$ を選択する.

これまで提案されてきた選択アルゴリズムとして代表的なものうち、(M1) にもとづくものとしては、Hoerl ら (1975) などがあり、(M2) にもとづくものとしては、Dempster ら (1977) などがある。ここでは後者の方法 (M2) にもとづいたアルゴリズムを考察する¹⁾。

本稿で扱う選択アルゴリズムに関する主なアイデアは、ある種の基準化を施した総平均 2 乗誤差 (相対効率と呼ぶ.) を利用することによって、推定すべきパラメータを一つのものに集約し、それを精度よく推定することによって、よりよいアルゴリズムを構築しようとするものである。

なお、本稿の構成としては、2 節でモデルと推定量について述べ、3 節で推定量の評価基準を与える。また、4 節でリッジ係数を選択するためのアルゴリズムについて考察し、5 節でモンテカルロ法によるシミュレーションの結果を与えたとともにそれらの考察を行う。最後に、6 節で総括を行う。

2 モデルと推定量

以下の (標準化) 線形回帰モデルを考える:

1) 方法 (M1) にもとづくものとして Hoerl ら (1975) の結果が保守的な結果を示すことがわかっている。

$$(M_X) \quad \mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

ここで、 \mathbf{Y} は n 次元応答変数ベクトルであり、 $\mathbf{X} = [\mathbf{1}, \mathbf{X}_p]$ は $n \times (p+1)$ 説明変数行列とする。ただし、 $\mathbf{1} = [1, \dots, 1]'$ ($\in \mathbb{R}^n$) とし、 $\mathbf{X}'_p \mathbf{1} = \mathbf{0}$ とする。ここで、 $\mathbf{0}$ は (p 次元) ゼロベクトルとし、 $'$ (プライム) は行列・ベクトルの転置を表す。また、 $\boldsymbol{\beta} = [\beta_0, \boldsymbol{\beta}'_p]'$ $= [\beta_0, \beta_1, \dots, \beta_p]'$ は ($p+1$) 次元回帰係数ベクトル (未知) である。さらに、 $\boldsymbol{\varepsilon}$ は n 次元誤差ベクトルであり、 n 次元正規分布 $N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ に従うものとする。なお、 $\mathbf{0}$ は (n 次元) ゼロベクトルであり、 $\mathbf{I}_n = \text{diag}(1, \dots, 1)$ は n 次元単位行列、 σ^2 は未知の誤差分散である。モデル (M_X) において切片項 β_0 とそれ以外の部分 $\boldsymbol{\beta}_p$ ($\in \mathbb{R}^p$: p 次元ユークリッド空間) に分離したときの表記:

$$\mathbf{Y} = \beta_0 \mathbf{1} + \mathbf{X}_p \boldsymbol{\beta}_p + \boldsymbol{\varepsilon}$$

も適宜利用される。モデル (M_X) が標準化モデル (standardized model) とよばれるのは、 \mathbf{X}_p の交差積行列 (情報行列) $\mathbf{X}'_p \mathbf{X}_p$ が相関形式 (correlation form) となるように標準化されていることによる。

回帰係数ベクトル $\boldsymbol{\beta}_p$ に対する通常最小自乗 (Ordinary Least Squares: OLS) 推定量は、正規方程式

$$\mathbf{X}'_p \mathbf{X}_p \boldsymbol{\beta}_p = \mathbf{X}'_p \mathbf{Y} \quad (1)$$

を $\boldsymbol{\beta}_p$ に関して解くことによって、

$$\widehat{\boldsymbol{\beta}}_p := (\mathbf{X}'_p \mathbf{X}_p)^{-1} \mathbf{X}'_p \mathbf{Y}$$

で与えられる。

標準化された説明変数行列 \mathbf{X}_p の列ベクトルの間に一次従属に近い関係が存在する場合、すなわち、多重共線性 (multicollinearity) が起こっている場合は OLS 推定量の推定精度が悪くなる場合があり、この問題に対して正規方程式 (1) の係数行列の対角成分に非負の値 k (リッジ係数 (ridge coefficient) とよばれる。) を加えることによって解を安定させることが Hoerl and Kennard (1970) によって提案された。この場合の (リッジ型) 正規方程式は、

$$(\mathbf{X}'_p \mathbf{X}_p + k \mathbf{I}_p) \boldsymbol{\beta}_p = \mathbf{X}'_p \mathbf{Y} \quad (2)$$

となり, この解を,

$$\widehat{\beta}_p(k) := (\mathbf{X}'_p \mathbf{X}_p + k \mathbf{I}_p)^{-1} \mathbf{X}'_p \mathbf{Y}$$

とおき, 通常リッジ回帰 (Ordinary Ridge Regression: ORR) 推定量とよぶ.

誤差分散 σ^2 の推定量としては, 以下のようなものが通常は利用される:

$$\widehat{\sigma}^2 := \frac{1}{n-p-1} (\mathbf{Y} - \mathbf{X}\widehat{\beta})' (\mathbf{Y} - \mathbf{X}\widehat{\beta}) \quad (3)$$

ここで, $\widehat{\beta} := (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ は回帰係数ベクトル β に関する OLS 推定量である. 以下のことが成り立つことに注意しよう:

$$E(\widehat{\sigma}^2) = \sigma^2 (\text{: 不偏性}), \quad \widehat{\sigma}^2 \sim \frac{\sigma^2}{n-p-1} \chi_{n-p-1}^2$$

ここで, E は推定量 (確率変数) の期待値を表す演算子であり, χ_{n-p-1}^2 は自由度 $n-p-1$ のカイ自乗分布を表す. また, “ \sim ” は確率変数がある種の分布に従うことを表す記号である.

3 評価基準

ORR 推定量の評価基準としては総平均 2 乗誤差 (Total Mean Square Error: TMSE) がしばしば利用される. OLS 推定量を例として TMSE の定義を以下に与える:

$$\begin{aligned} \text{TMSE}(\widehat{\beta}_p) &:= E(\widehat{\beta}_p - \beta_p)' (\widehat{\beta}_p - \beta_p) \\ &= \text{trace} V(\widehat{\beta}_p) + \text{bias}(\widehat{\beta}_p)' \text{bias}(\widehat{\beta}_p) \end{aligned}$$

ここで, V は (推定量の) 分散共分散行列を求めるための演算子を表し, trace は行列のトレース²⁾(trace), また bias は推定量の偏り (bias) を表す. 実際には,

$$\begin{aligned} E(\widehat{\beta}_p) &= (\mathbf{X}'_p \mathbf{X}_p)^{-1} \mathbf{X}'_p E(\mathbf{Y}) = \beta_p \\ V(\widehat{\beta}_p) &:= E(\widehat{\beta}_p - E(\widehat{\beta}_p)) (\widehat{\beta}_p - E(\widehat{\beta}_p))' \\ &= E(\widehat{\beta}_p - \beta_p) (\widehat{\beta}_p - \beta_p)' = \sigma^2 (\mathbf{X}'_p \mathbf{X}_p)^{-1} \end{aligned}$$

より,

2) 対角成分の和

$$\begin{aligned}\text{traceV}(\widehat{\beta}_p) &= \sigma^2 \text{trace}(\mathbf{X}'_p \mathbf{X}_p)^{-1} \\ \text{bias}(\widehat{\beta}_p) &:= \text{E}(\widehat{\beta}_p) - \beta_p = \mathbf{0}\end{aligned}$$

となり，結局，

$$\text{TMSE}(\widehat{\beta}_p) = \sigma^2 \text{trace}(\mathbf{X}'_p \mathbf{X}_p)^{-1} = \sigma^2 \sum_{i=1}^p \frac{1}{\lambda_i}$$

となる．ここで， $\lambda_1 \geq \dots \geq \lambda_p (> 0)$ は情報行列 $\mathbf{X}'_p \mathbf{X}_p$ の固有値であり，

$$\mathbf{\Lambda}_p := \text{diag}(\lambda_1, \dots, \lambda_p)$$

とし， $\mathbf{\Gamma}_p$ を行列 $\mathbf{X}'_p \mathbf{X}_p$ を対角化する p 次の直交行列 ($\mathbf{\Gamma}'_p \mathbf{\Gamma}_p = \mathbf{\Gamma}_p \mathbf{\Gamma}'_p = \mathbf{I}_p$) とすると，

$$\mathbf{X}'_p \mathbf{X}_p = \mathbf{\Gamma}_p \mathbf{\Lambda}_p \mathbf{\Gamma}'_p; \text{ スペクトル分解}$$

が成り立つことから，

$$\text{trace}(\mathbf{X}'_p \mathbf{X}_p)^{-1} = \text{trace} \mathbf{\Gamma}_p \mathbf{\Lambda}_p^{-1} \mathbf{\Gamma}'_p = \text{trace} \mathbf{\Lambda}_p^{-1} \mathbf{\Gamma}'_p \mathbf{\Gamma}_p = \text{trace} \mathbf{\Lambda}_p^{-1} = \sum_{i=1}^p \frac{1}{\lambda_i}$$

となることを使った．

次に，ORR 推定量に対する TMSE を求める．まず，

$$\begin{aligned}\text{TMSE}(\widehat{\beta}_p(k)) &= \text{E}(\widehat{\beta}_p(k) - \beta_p)' (\widehat{\beta}_p(k) - \beta_p) \\ &= \text{traceV}(\widehat{\beta}_p(k)) + \text{bias}(\widehat{\beta}_p(k))' \text{bias}(\widehat{\beta}_p(k))\end{aligned}$$

であり，ここで，

$$\begin{aligned}\text{E}(\widehat{\beta}_p(k)) &= (\mathbf{X}'_p \mathbf{X}_p + k \mathbf{I}_p)^{-1} \mathbf{X}'_p \text{E}(Y) = (\mathbf{X}'_p \mathbf{X}_p + k \mathbf{I}_p)^{-1} \mathbf{X}'_p \mathbf{X}_p \beta_p \\ \text{bias}(\widehat{\beta}_p(k)) &= \text{E}(\widehat{\beta}_p(k)) - \beta_p = \{(\mathbf{X}'_p \mathbf{X}_p + k \mathbf{I}_p)^{-1} \mathbf{X}'_p \mathbf{X}_p - \mathbf{I}_p\} \beta_p \\ \text{V}(\widehat{\beta}_p(k)) &= \text{E}(\widehat{\beta}_p(k) - \text{E}(\widehat{\beta}_p(k))) (\widehat{\beta}_p(k) - \text{E}(\widehat{\beta}_p(k)))' \\ &= \sigma^2 (\mathbf{X}'_p \mathbf{X}_p + k \mathbf{I}_p)^{-1} \mathbf{X}'_p \mathbf{X}_p (\mathbf{X}'_p \mathbf{X}_p + k \mathbf{I}_p)^{-1}\end{aligned}$$

であることから，

$$\begin{aligned}\text{TMSE}(\widehat{\beta}_p(k)) &= \sigma^2 \text{trace}(\mathbf{X}'_p \mathbf{X}_p + k \mathbf{I}_p)^{-1} \mathbf{X}'_p \mathbf{X}_p (\mathbf{X}'_p \mathbf{X}_p + k \mathbf{I}_p)^{-1} \\ &\quad + \beta'_p \{ \mathbf{X}'_p \mathbf{X}_p (\mathbf{X}'_p \mathbf{X}_p + k \mathbf{I}_p)^{-1} - \mathbf{I}_p \} \{ (\mathbf{X}'_p \mathbf{X}_p + k \mathbf{I}_p)^{-1} \mathbf{X}'_p \mathbf{X}_p - \mathbf{I}_p \} \beta_p\end{aligned}$$

となる．さらに，

$$\boldsymbol{\alpha}_p := \boldsymbol{\Gamma}_p' \boldsymbol{\beta}_p$$

とおくことによって、 $\boldsymbol{\Gamma}_p$ の直交性に注意すると、

$$\begin{aligned} \text{TMSE}(\widehat{\boldsymbol{\beta}}_p(k)) &= \sigma^2 \text{trace} \boldsymbol{\Gamma}_p (\boldsymbol{\Lambda}_p + k \mathbf{I}_p)^{-1} \boldsymbol{\Lambda}_p (\boldsymbol{\Lambda}_p + k \mathbf{I}_p)^{-1} \boldsymbol{\Gamma}_p' \\ &\quad + \boldsymbol{\beta}_p' \boldsymbol{\Gamma}_p \{ \boldsymbol{\Lambda}_p (\boldsymbol{\Lambda}_p + k \mathbf{I}_p)^{-1} - \mathbf{I}_p \} \{ (\boldsymbol{\Lambda}_p + k \mathbf{I}_p)^{-1} \boldsymbol{\Lambda}_p - \mathbf{I}_p \} \boldsymbol{\Gamma}_p' \boldsymbol{\beta}_p \\ &= \sigma^2 \text{trace} (\boldsymbol{\Lambda}_p + k \mathbf{I}_p)^{-1} \boldsymbol{\Lambda}_p (\boldsymbol{\Lambda}_p + k \mathbf{I}_p)^{-1} \\ &\quad + \boldsymbol{\alpha}_p' \{ \boldsymbol{\Lambda}_p (\boldsymbol{\Lambda}_p + k \mathbf{I}_p)^{-1} - \mathbf{I}_p \} \{ (\boldsymbol{\Lambda}_p + k \mathbf{I}_p)^{-1} \boldsymbol{\Lambda}_p - \mathbf{I}_p \} \boldsymbol{\alpha}_p \\ &= \sigma^2 \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + k)^2} + \sum_{i=1}^p \frac{\alpha_i^2 k^2}{(\lambda_i + k)^2} =: \rho(k) \end{aligned}$$

を得る。

さて、ORR 推定量において $k = 0$ のとき、OLS 推定量となること、すなわち、

$$\widehat{\boldsymbol{\beta}}_p(0) = \widehat{\boldsymbol{\beta}}_p = (\mathbf{X}_p' \mathbf{X}_p)^{-1} \mathbf{X}_p' \mathbf{Y}$$

という事実は、それらの TMSE の間にも当然成り立ち、

$$\text{TMSE}(\widehat{\boldsymbol{\beta}}_p(0)) = \rho(0) = \sigma^2 \sum_{i=1}^p \frac{1}{\lambda_i} = \text{TMSE}(\widehat{\boldsymbol{\beta}}_p)$$

となることに注意しよう。

注意 3.1 (正準形式) ORR 推定量を TMSE $\rho(k)$ で評価する限り、情報行列 $\mathbf{X}_p' \mathbf{X}_p$ をスペクトル分解したときに得られる固有値 λ_i と、回帰係数ベクトル $\boldsymbol{\beta}_p$ を直交行列 $\boldsymbol{\Gamma}_p$ を用いて変換した $\boldsymbol{\alpha}_p = \boldsymbol{\Gamma}_p' \boldsymbol{\beta}_p$ の成分 α_i が主に扱われる。このことから、便宜上、

$$(M_A) \quad \mathbf{Y} = \mathbf{A} \boldsymbol{\alpha} + \boldsymbol{\varepsilon} = \alpha_0 \mathbf{1} + \mathbf{A}_p \boldsymbol{\alpha}_p + \boldsymbol{\varepsilon}$$

とあらかじめ変換しておいたモデルが利用される場合がある。このモデルを正準形式 (canonical form) という。ここで、

$$\mathbf{A} := \mathbf{X} \boldsymbol{\Gamma}, \quad \mathbf{A}_p := \mathbf{X}_p \boldsymbol{\Gamma}_p, \quad \boldsymbol{\alpha} := \boldsymbol{\Gamma}' \boldsymbol{\beta}, \quad \boldsymbol{\alpha}_p := \boldsymbol{\Gamma}_p' \boldsymbol{\beta}_p$$

であり、

$$\boldsymbol{\Gamma} := \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_p \end{bmatrix}, \quad \boldsymbol{\alpha} := \begin{bmatrix} \alpha_0 \\ \boldsymbol{\alpha}_p \end{bmatrix}$$

とおいた。正準形式 (M_A) における回帰係数ベクトル α_p の OLS 推定量 $\hat{\alpha}_p$ と ORR 推定量 $\hat{\alpha}_p(k)$ は, それぞれ,

$$\hat{\alpha}_p := (\mathbf{A}'_p \mathbf{A}_p)^{-1} \mathbf{A}'_p \mathbf{Y} = \Lambda_p^{-1} \mathbf{A}'_p \mathbf{Y},$$

$$\hat{\alpha}_p(k) := (\mathbf{A}'_p \mathbf{A}_p + k \mathbf{I}_p)^{-1} \mathbf{A}'_p \mathbf{Y} = (\Lambda_p + k \mathbf{I}_p)^{-1} \mathbf{A}'_p \mathbf{Y},$$

で定義される。ここで, OLS 推定量と ORR 推定量の間には,

$$\hat{\alpha}_p(k) = (\mathbf{I}_p + k \Lambda_p^{-1})^{-1} \hat{\alpha}_p \iff \hat{\alpha}_i(k) = \frac{\lambda_i}{\lambda_i + k} \hat{\alpha}_i, \quad i = 1, \dots, p \quad (4)$$

なる関係が成り立つことに注意しよう。また,

$$\hat{\alpha}_p \sim N_p(\alpha_p, \sigma^2 \Lambda_p^{-1}),$$

$$\hat{\alpha}_p(k) \sim N_p((\mathbf{I}_p + k \Lambda_p^{-1})^{-1} \alpha_p, \sigma^2 (\mathbf{I}_p + k \Lambda_p^{-1})^{-1} \Lambda_p^{-1} (\mathbf{I}_p + k \Lambda_p^{-1})^{-1})$$

が成り立つことに注意しよう。

標準化モデルと正準形式に対する OLS 推定量と ORR 推定量の間の

$$\hat{\alpha}_p = \Gamma'_p \hat{\beta}_p,$$

$$\hat{\alpha}_p(k) = \Gamma'_p \hat{\beta}_p(k)$$

なる関係を使うと, これらの推定量の TMSE は,

$$\text{TMSE}(\hat{\alpha}_p) = \text{TMSE}(\hat{\beta}_p) = \rho(0),$$

$$\text{TMSE}(\hat{\alpha}_p(k)) = \text{TMSE}(\hat{\beta}_p(k)) = \rho(k)$$

となり, TMSE で評価する限りにおいては, 正準形式 (M_A) のもとで考えても一般性は失わないことに注意しよう。

4 選択アルゴリズム

4.1 相対効率

本稿におけるアイデアは TMSE を基準化したものを利用することによって, 推定すべきパラメータを非集中度に特定化し, それを精度よく推定しようとするものである。すなわち, リッジ係数 k を選択するために直接 TMSE $\rho(k)$ を利用するのではなく, OLS 推定量の TMSE $\rho(0)$ で除したものを新たな評価基

準として定義する³⁾:

$$\mathcal{R}(k) := \frac{\rho(k)}{\rho(0)} = \sum_{i=1}^p w_i \frac{\lambda_i^2}{(\lambda_i + k)^2} + \sum_{i=1}^p w_i \delta_i \frac{k^2}{(\lambda_i + k)^2} = \sum_{i=1}^p \frac{w_i(\lambda_i^2 + \delta_i k^2)}{(\lambda_i + k)^2} \quad (5)$$

ここで、 w_i は以下で定義される固有値にもとづくある種の重みである:

$$w_i := \frac{1/\lambda_i}{\sum_{j=1}^p 1/\lambda_j} (> 0), \quad \sum_{i=1}^n w_i = 1$$

また、 δ_i は以下で定義されるパラメータ (非心度) である:

$$\delta_i := \frac{\alpha_i^2}{\sigma^2/\lambda_i}$$

本稿では、この評価基準 $\mathcal{R}(k)$ を相対効率 (Relative Efficiency: REff) と呼ぶことにする⁴⁾.

以下の命題は相対効率の形状を把握する上で重要である:

命題 4.1 相対効率 $\mathcal{R}(k)$ に関して以下のことが成り立つ:

(i) $\mathcal{R}(k) > 0, \quad \mathcal{R}(0) = 1$

(ii) $\lim_{k \rightarrow \infty} \mathcal{R}(k) = \frac{\sum_{i=1}^p \alpha_i^2}{\sigma^2 / \sum_{i=1}^p (1/\lambda_i)} \left(= \frac{\|E(\hat{\alpha}_p)\|^2}{\text{traceV}(\hat{\alpha}_p)} = \frac{\|E(\hat{\beta}_p)\|^2}{\text{traceV}(\hat{\beta}_p)} \right)$

(iii) 任意の $k \in K_1$ に対して、 $\mathcal{R}(k)$ は減少し、任意の $k \in K_2$ に対して $\mathcal{R}(k)$ は凸である.

(iv) 任意の $k \in K_1$ に対して、 $\mathcal{R}(k)$ は減少かつ凸である.

ここで、

$$K_1 := \left(0, \min_{i \in N_p} \frac{\lambda_i}{\delta_i} \right), \quad K_2 := \left(0, \min_{i \in N_p} \left\{ \frac{1}{2} \left(\lambda_i + \frac{\lambda_i}{\delta_i} \right) + \frac{\lambda_i}{\delta_i} \right\} \right)$$

である.

3) この評価基準自体は、新しいものではなく、これまでも利用されていることに注意しよう. (たとえば、Nordberg (1982) の 4 節参照.)

4) ここで、このように呼ぶ理由としては、ORR 推定量の OLS 推定量に対する相対的な効率であるためである.

証明: (i) は明らか. (ii) は,

$$\lim_{k \rightarrow \infty} \frac{\lambda_i^2}{(\lambda_i + k)^2} = 0, \quad \lim_{k \rightarrow \infty} \frac{k^2}{(\lambda_i + k)^2} = 1$$

に注意し, δ_i と w_i の定義を使うと,

$$\begin{aligned} \lim_{k \rightarrow \infty} \mathcal{R}(k) &= \sum_{i=1}^p w_i \left(\lim_{k \rightarrow \infty} \frac{\lambda_i^2}{(\lambda_i + k)^2} \right) + \sum_{i=1}^p w_i \delta_i \left(\lim_{k \rightarrow \infty} \frac{k^2}{(\lambda_i + k)^2} \right) \\ &= \sum_{i=1}^p w_i \delta_i = \frac{\sum_{i=1}^p \alpha_i^2}{\sigma^2 / \sum_{i=1}^p (1/\lambda_i)} \end{aligned}$$

となり, 直接示すことができる.

次に, (iii) を示す. 1, 2 次微分は,

$$\frac{d}{dk} \mathcal{R}(k) = \sum_{i=1}^p \frac{2w_i \lambda_i (\delta_i k - \lambda_i)}{(\lambda_i + k)^3}, \quad (6)$$

$$\frac{d^2}{dk^2} \mathcal{R}(k) = - \sum_{i=1}^p \frac{2w_i \lambda_i (2\delta_i k - \delta_i \lambda_i - 3\lambda_i)}{(\lambda_i + k)^4} \quad (7)$$

となる. これらの結果から, $N_p := \{1, \dots, p\}$ とおくと,

$$\forall i \in N_p, \delta_i k - \lambda_i < 0 \iff (0 <) k < \min_{i \in N_p} \frac{\lambda_i}{\delta_i}$$

となり, 少なくとも,

$$\forall k \in \left(0, \min_{i \in N_p} \frac{\lambda_i}{\delta_i} \right) = K_1$$

という範囲では,

$$\frac{d}{dk} \mathcal{R}(k) < 0$$

となり, 減少していることがわかる. また,

$$\forall i \in N_p, 2\delta_i k - \delta_i \lambda_i - 3\lambda_i < 0 \iff (0 <) k < \min_{i \in N_p} \left\{ \frac{1}{2} \left(\lambda_i + \frac{\lambda_i}{\delta_i} \right) + \frac{\lambda_i}{\delta_i} \right\}$$

であることから,

$$\forall k \in \left(0, \min_{i \in N_p} \left\{ \frac{1}{2} \left(\lambda_i + \frac{\lambda_i}{\delta_i} \right) + \frac{\lambda_i}{\delta_i} \right\} \right) = K_2$$

という範囲では,

$$\frac{d^2}{dk^2} \mathcal{R}(k) > 0$$

となり, 凸であることがわかる.

最後に, (iv) は,

$$\forall i \in N_p, \frac{\lambda_i}{\delta_i} < \frac{1}{2} \left(\lambda_i + \frac{\lambda_i}{\delta_i} \right) + \frac{\lambda_i}{\delta_i}$$

であることから, 上記の 2 つの区間の間に $K_1 \subset K_2$ が成り立ち, 結局,

$$\forall k \in K_1, \frac{d}{dk} \mathcal{R}(k) < 0 \text{ かつ } \frac{d^2}{dk^2} \mathcal{R}(k) > 0$$

が成り立つ. □

図 1, 2, 3, 4 に Hoerl (1962) によって与えられた 2 種類のデータセット (DS1), (DS2) に対する TMSE と相対効率 (REff) に関するプロットを与える. (実線が TMSE または相対効率 (REff) を表し, 破線が分散部分, 点線が偏り部分を表すことに注意しよう.) なお, Hoerl (1962) のデータセットについては 4.3 節にも簡単な説明を与えているので参照されたい.

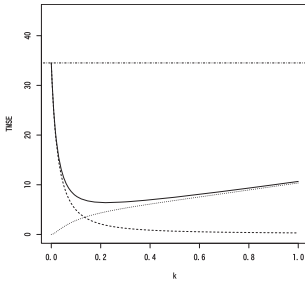


図 1: Hoerl のデータセット (DS1) に対する TMSE

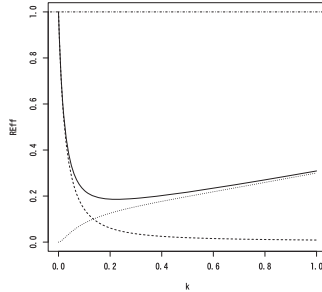


図 2: Hoerl のデータセット (DS1) に対する相対効率 (REff)

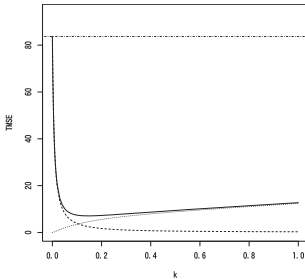


図 3: Hoerl のデータセット (DS2) に対する TMSE

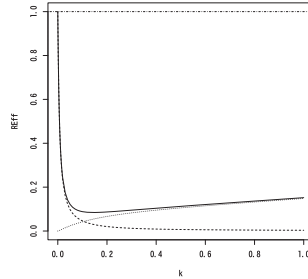


図 4: Hoerl のデータセット (DS2) に対する相対効率 (REff)

注意 4.1 TMSE $\rho(k)$ と相対効率 $\mathcal{R}(k)$ の最小値を与える $k = k^*$ は一致することに注意しよう⁵⁾:

$$\arg \min_{k>0} \rho(k) = \arg \min_{k>0} \mathcal{R}(k) = k^*$$

ここで, $\arg \min_{x \in D} f(x)$ は, 領域 D 内で関数 $f(x)$ を最小にする x の値を表す記号であることに注意しよう.

注意 4.2 (非心度) $\delta_i = \alpha_i^2 / (\sigma^2 / \lambda_i)$ が非心度 (non-centrality parameter) になっていることは以下のような理由による. まず,

$$\hat{\alpha}_i \sim N\left(\alpha_i, \frac{\sigma^2}{\lambda_i}\right), \quad \frac{\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-p-1}^2 / (n-p-1), \quad \hat{\alpha}_i \perp \hat{\sigma}^2$$

であることと,

$$\frac{\hat{\alpha}_i}{\sigma / \sqrt{\lambda_i}} \sim N\left(\frac{\alpha_i}{\sigma / \sqrt{\lambda_i}}, 1\right) \implies \frac{\hat{\alpha}_i^2}{\sigma^2 / \lambda_i} \sim \chi_1^2(\delta_i)$$

($\chi_1^2(\delta_i)$; 自由度 1, 非心度 δ_i の非心カイ自乗分布)

となることから,

$$\frac{\hat{\alpha}_i^2}{\hat{\sigma}^2 / \lambda_i} = \frac{\hat{\alpha}_i^2}{\sigma^2 / \lambda_i} \sim \frac{\chi_1^2(\delta_i) / 1}{\chi_{n-p-1}^2 / (n-p-1)} = F_{n-p-1}^1(\delta_i)$$

($F_{n-p-1}^1(\delta_i)$; 自由度 $(1, n-p-1)$, 非心度 δ_i の非心エフ分布)

となるからである. なお, 非心カイ自乗分布と非心エフ分布に関しては付録 A を参照されたい.

4.2 相対効率の推定

理想的なアルゴリズムとしては, 相対効率 $\mathcal{R}(k)$ を最小にする k を求めることができればよいけれども, 非心度 δ_i という未知のパラメータに依存することから現実的には困難である. よって, 実際には相対効率を推定する必要がある. 相対効率の通常の推定量として以下のものが提案される:

$$\hat{\mathcal{R}}(k) := \sum_{i=1}^p w_i \frac{\lambda_i^2}{(\lambda_i + k)^2} + \sum_{i=1}^p w_i \hat{\delta}_i \frac{k^2}{(\lambda_i + k)^2} = \sum_{i=1}^p \frac{w_i (\lambda_i^2 + \hat{\delta}_i k^2)}{(\lambda_i + k)^2} \quad (8)$$

5) TMSE $\rho(k)$ は, パラメータの値の取り方によっては, $k > 0$ のすべての値に関して凸ではないことが知られており, その場合は原点にもっとも近い極小値を k^* と考える.

ここで、

$$\widehat{\delta}_i := \frac{\widehat{\alpha}_i^2}{\widehat{\sigma}^2/\lambda_i}, \quad i = 1, \dots, p \quad (9)$$

であり、これは非心度 δ_i の通常の推定量と考えることができる。

一方、相対効率の他の推定量として以下のものが提案される：

$$\overline{\mathcal{R}}(k) := \sum_{i=1}^p w_i \frac{\lambda_i^2}{(\lambda_i + k)^2} + \sum_{i=1}^p w_i \widehat{\delta}_i \frac{k^2}{(\lambda_i + k)^2} = \sum_{i=1}^p \frac{w_i(\lambda_i^2 + \widehat{\delta}_i k^2)}{(\lambda_i + k)^2} \quad (10)$$

ここで、

$$\overline{\delta}_i := \frac{n-p-3}{n-p-1} \widehat{\delta}_i - 1 = \frac{n-p-3}{n-p-1} \frac{\widehat{\alpha}_i^2}{\widehat{\sigma}^2/\lambda_i} - 1, \quad i = 1, \dots, p \quad (11)$$

であり⁶⁾、この推定量は非心度 δ_i の不偏推定量である⁷⁾：

$$E(\overline{\delta}_i) = \delta_i, \quad i = 1, \dots, p$$

また、非心度 δ_i における回帰係数の推定に ORR 推定量を使った、

$$\widehat{\delta}_i := \frac{\widehat{\alpha}_i^2(k)}{\widehat{\sigma}^2/\lambda_i} = \frac{\lambda_i^2}{(\lambda_i + k)^2} \widehat{\delta}_i, \quad i = 1, \dots, p \quad (12)$$

も考えることができる。ここで、 $\widehat{\alpha}_i(k)$ は正準形式における回帰係数 α_i に対する ORR 推定量 ((4) 式参照.) であり、 $\widehat{\alpha}_i^2(k) = (\widehat{\alpha}_i(k))^2$ に注意しよう。この推定量を利用した相対効率の推定量は、

$$\widehat{\mathcal{R}}(k) := \sum_{i=1}^p w_i \frac{\lambda_i^2}{(\lambda_i + k)^2} + \sum_{i=1}^p w_i \widehat{\delta}_i \frac{k^2}{(\lambda_i + k)^2} = \sum_{i=1}^p \frac{w_i(\lambda_i^2 + \widehat{\delta}_i k^2)}{(\lambda_i + k)^2} \quad (13)$$

である。

さらに、 $\overline{\delta}_i$ と $\widehat{\delta}_i$ を折衷した以下のような推定量も考えることができる：

$$\overline{\overline{\delta}}_i := \frac{\lambda_i^2}{(\lambda_i + k)^2} \overline{\delta}_i, \quad i = 1, \dots, p \quad (14)$$

この推定量を利用した相対効率の推定量は、

$$\overline{\overline{\mathcal{R}}}(k) := \sum_{i=1}^p w_i \frac{\lambda_i^2}{(\lambda_i + k)^2} + \sum_{i=1}^p w_i \overline{\overline{\delta}}_i \frac{k^2}{(\lambda_i + k)^2} = \sum_{i=1}^p \frac{w_i(\lambda_i^2 + \overline{\overline{\delta}}_i k^2)}{(\lambda_i + k)^2} \quad (15)$$

である。

6) ここで、自由度を $(\nu_1, \nu_2) := (1, n-p-1)$ とおくと、

$$\overline{\delta}_i \sim \frac{\nu_1(\nu_2 - 2)}{\nu_2} F_{\nu_2}^{\nu_1}(\delta_i) - \nu_1$$

と書くことができることに注意しよう。

7) さらに、最小分散不偏推定量 (Uniform Minimum Variance Unbiased Estimator: UMVUE) となっていることにも注意しよう。(Perlman and Rasmussen (1975), Johnson, Kotz, and Balakrishnan (1995) 参照.)

注意 4.3 本稿の文脈から、相対効率 $\mathcal{R}(k)$ の推定を考えるとすることは、非心エフ分布の非心度 δ_i を推定することと同等であることが理解される。ここで与えたように、様々なタイプの非心度の推定量を考えることが必要となる理由は、自由度と非心度に依存するけれども、非心エフ分布が右裾が重くなる (heavy right-tail) 傾向にあり、(9) 式で定義された通常の推定量 $\hat{\delta}_i$ (非心エフ分布の統計量そのものであることに注意) が非心度の真の値を大きく過大評価 (over estimate) することによる。このことは、自由度 $(\nu_1, \nu_2) = (1, 6)$ 、非心度 $\delta = 200$ の非心エフ分布 $F_{\nu_2}^{\nu_1}(\delta)$ に従うデータ (疑似乱数) に関するヒストグラムとボックスプロットからも容易に理解できるであろう⁸⁾。(図 5 参照。) 以上のことから、過大評価をおさえるような推定量 (ある意味での縮小推定量) の必要性が示唆され、(11), (12), (14) 式で定義された推定量 $\bar{\delta}_i, \hat{\delta}_i, \bar{\delta}_i$ はいずれもそのような修正が施されていることに注意しよう。

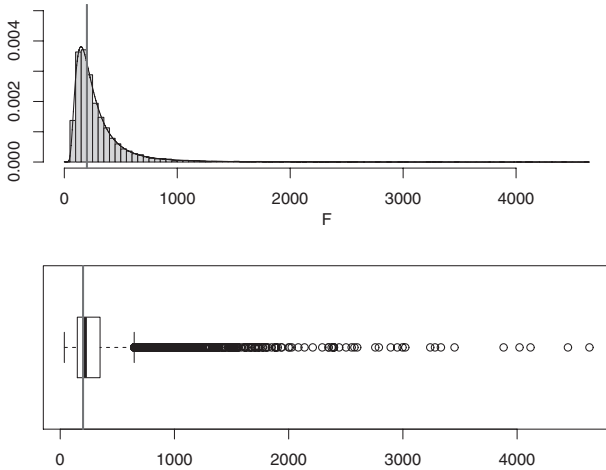


図 5: 自由度 $(\nu_1, \nu_2) = (1, 6)$ 、非心度 $\delta = 200$ の非心エフ分布からの 10000 個のデータに関する (密度関数付き) ヒストグラムとボックスプロット (グラフ中の「垂直線」は非心度の真値を表す。)

8) 非心度の真値が 200 であるにも関わらず、その 10 倍の 2000 を超えるデータ (推定値) がでる可能性があることを示している。

注意 4.4 非心度の推定量 $\bar{\delta}_i$ は負の値をとる可能性がある. たとえば, $n = 10$, $p = 3$, $\delta_i = 0.02$ ⁹⁾ のとき, $\widehat{\delta}_i(> 0) \sim F_6^1(0.02)$ に注意すると,

$$\begin{aligned} P(\bar{\delta}_i < 0) &= P\left(0 < \widehat{\delta}_i < \frac{n-p-1}{n-p-3}\right) = P\left(0 < \widehat{\delta}_i < \frac{3}{2}\right) \\ &= 0.7291485 \end{aligned}$$

となり, 約 72% の確率で負の値が出現することを示している. このことから, $\delta_i \geq 0$ という仮定と不整合であるため,

$$\widetilde{\delta}_i := \max(0, \bar{\delta}_i), \quad i = 1, \dots, p \quad (16)$$

という推定量を利用し,

$$\widetilde{\mathcal{R}}(k) := \sum_{i=1}^p w_i \frac{\lambda_i^2}{(\lambda_i + k)^2} + \sum_{i=1}^p w_i \widetilde{\delta}_i \frac{k^2}{(\lambda_i + k)^2} \quad (17)$$

なる相対効率の推定量を構成することが示唆される. しかしながら, $\bar{\delta}_i = 0$ が高頻度で現れると (17) 式の右辺第二項 (バイアス部分) のうちの一部が 0 となり, 関数の構造 (形状) が大きく影響を受けることになる. 次項で扱うが, 我々が知りたい「情報」は, 相対効率の推定量の最小値を与える k の値であることから, 推定量 $\widetilde{\delta}_i$ を用いた場合は本稿では割愛する¹⁰⁾.

4.3 リッジ係数の選択アルゴリズム

本稿におけるリッジ係数の選択アルゴリズムは, 相対効率 $\mathcal{R}(k)$ の推定量を最小にする k の値を求めることによって与えられる. たとえば, (8) 式によって定義された相対効率の推定量 $\widehat{\mathcal{R}}(k)$ の場合は,

$$\widehat{k} := \arg \min_{k>0} \widehat{\mathcal{R}}(k)$$

を求めることによってリッジ係数 k を選択する. なお, 最小値を与える k を求める具体的なアルゴリズムは, 相対効率 $\mathcal{R}(k)$ の 1 次導関数の零点を Newton-Raphson 法によって求めるというものである. すなわち, 初期値 k_0 を適切に与え,

9) Hoerl(1962) のデータセット (DS2) の場合の非心度 δ_3 に対応する.

10) 実際に, 事前に行ったシミュレーションにおいてもよい結果を与えることはなかった.

$$k_{m+1} := k_m - \frac{\widehat{\mathcal{R}}'(k_m)}{\widehat{\mathcal{R}}''(k_m)}, \quad m = 0, 1, 2, \dots$$

を繰り返し計算し、あらかじめ決められた $\epsilon (> 0)$ に対して、

$$|k_{m+1} - k_m| < \epsilon$$

となったとき、 $\widehat{k} = k_{m+1}$ とする方法である。ここで、 $\widehat{\mathcal{R}}'(k_m)$ 、 $\widehat{\mathcal{R}}''(k_m)$ は、それぞれ、 $k = k_m$ おける相対効率の推定量 $\widehat{\mathcal{R}}(k)$ の 1, 2 次導関数の値である。なお、本稿を通じて、初期値は $k_0 = 0$ を与えて繰り返し行っており、もし最終的に選択されたリッジ係数が、負になった場合は 0 としていることに注意しよう。

注意 4.5 Lee and Campbell (1985) は、TMSE $\rho(k)$ の推定量:

$$\widehat{\rho}(k) := \widehat{\sigma}^2 \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + k)^2} + \sum_{i=1}^p \frac{\widehat{\alpha}_i^2 k^2}{(\lambda_i + k)^2}$$

を利用し Newton-Raphson 法を利用して最小値を与える k を求めるアルゴリズムを提案している。この方法は相対効率の推定量 $\widehat{\mathcal{R}}(k)$ にもとづいて選択される \widehat{k} と本質的に同じものであり、 $\widehat{\mathcal{R}}(k)$ にもとづいて選択される \widehat{k} はすでに提案されたものであることに注意しよう。

また、Jimichi and Inagaki (1993) は、TMSE $\rho(k)$ の推定量として、

$$\bar{\rho}(k) := \widehat{\sigma}^2 \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + k)^2} + \sum_{i=1}^p \frac{\bar{\alpha}_i^2 k^2}{(\lambda_i + k)^2}$$

を提案し、Newton-Raphson 法を利用して最小値を与える k を求めるアルゴリズムを提案している。ここで、

$$\bar{\alpha}_i^2 := \widehat{\alpha}_i^2 - \frac{\widehat{\sigma}^2}{\lambda_i}$$

であり、これは α_i^2 の不偏推定量:

$$E(\bar{\alpha}_i^2) := \alpha_i^2$$

となっている。これより、 $\bar{\rho}(k)$ は TMSE の不偏推定量ともなっていることに注意しよう:

$$E(\bar{\rho}(k)) = \rho(k)$$

この方法は、相対効率の推定量 $\overline{\mathcal{R}}(k)$ を利用したアルゴリズム \overline{k} と類似のものであるが、厳密には異なることに注意しよう。

さらに、Dempster ら (1977) は、適切に選択された k の値に対して、

$$\left| \frac{d\widehat{\rho}(k)}{dk} \right| = \left| 2 \sum_{i=1}^p \frac{\lambda_i (\widehat{\alpha}_i^2(k) k - \widehat{\sigma}^2)}{(\lambda_i + k)^3} \right|$$

を評価し、最小値を与える k を選択することによって、TMSE $\rho(k)$ の 1 次導関数の零点を推定するアルゴリズムを提案した¹¹⁾。この方法は、本稿で扱う $\widehat{\mathcal{R}}(k)$ 、 $\overline{\mathcal{R}}(k)$ を利用したアルゴリズムと類似したものであるが、Newton-Raphson 法を利用していないという点で決定的に異なっている。つまり、Dempster ら (1977) のアルゴリズムは、前もって k に関する候補点を選んでおく必要があるのに対して、本稿で扱われているアルゴリズムは初期値さえ与えれば自動的に k が選択されることに注意しよう。なお、Dempster ら (1977) のアルゴリズムの性能は、候補点の選び方に強く依存することに注意が必要である¹²⁾。

注意 4.6 相対効率の推定量 $\widehat{\mathcal{R}}(k)$ 、 $\overline{\mathcal{R}}(k)$ に対する導関数は相対効率そのものの導関数 (6), (7) と本質的にかわらないけれども、相対効率の推定量 $\widehat{\mathcal{R}}(k)$ 、 $\overline{\mathcal{R}}(k)$ の導関数は非心度の推定量 $\widehat{\delta}_i$ 、 $\overline{\delta}_i$ が、それぞれ、 k に依存するため異なることに注意しよう。具体的な 1, 2 次導関数については付録 B を参照されたい。

5 モンテカルロ・シミュレーション

この節では、本稿で与えられたリッジ係数の選択アルゴリズムに関するモンテカルロ法によるシミュレーションを与える。なお、シミュレーションを実行した計算機環境については付録 C を参照されたい。

5.1 データセット

データセットとしては、Hoerl (1962) によって与えられた 2 種類のものを利用した。これらのデータセットの詳細は、典拠である Hoerl (1962) を参照す

11) Dempster ら (1977) による方法はメッシュ法 (mesh method) の一種といえる。

12) 言い換えると、候補点の選び方が別途問題となる。

るか、またはこのデータを利用した種々のモンテカルロ・シミュレーションが、たとえば、地道 (2002), 地道 (2004), Jimichi (2005) などに与えられているので参照されたい。それぞれのデータセットに対するパラメータの真値に関する主要量とそれらの推定値が表 1, 2 に与えられている。なお、両方のデータセットに対して、 $n = 10$, $p = 3$, $n - p - 1 = 6$, $\sigma^2 = 1$ であることに注意しよう。(便宜上、パラメータの添字には j を使用していることに注意しよう。)

表 1: データセット (DS1) に対するパラメータと固有値に関する主要な値

j	α_j	α_j^2	λ_j	σ^2/λ_j	w_j	δ_j
1	-8.04	64.61	2.87	0.35	0.01	185.23
2	1.63	2.67	0.09	11.14	0.32	0.24
3	-2.03	4.12	0.04	23.03	0.67	0.18

表 2: データセット (DS2) に対するパラメータと固有値に関する主要な値

j	α_j	α_j^2	λ_j	σ^2/λ_j	w_j	δ_j
1	-7.37	54.37	2.89	0.35	0.00	157.01
2	2.97	8.83	0.10	10.15	0.12	0.87
3	-1.25	1.57	0.01	73.19	0.87	0.02

5.2 実行手順

モンテカルロ・シミュレーションの実行プロセスは以下の手順で行った。ただし、ここではリッジ係数が $k = \hat{k}$ の場合について説明するが、 $k = \bar{k}, \underline{k}, \overline{k}$ についても同様であることに注意しよう。なお、シミュレーションの繰り返し回数 $N = 10^6$ とした。

(Step1) 乱数の生成:

$$\hat{\alpha}_{ij} \stackrel{\text{i.i.d.}}{\sim} N\left(\alpha_j, \frac{\sigma^2}{\lambda_j}\right), \quad \hat{\sigma}_i^2 \stackrel{\text{i.i.d.}}{\sim} \frac{\sigma^2}{n-p-1} \chi_{n-p-1}^2,$$

$$i = 1, \dots, N, \quad j = 1, \dots, p$$

(Step2) 非心度の推定:

$$\hat{\delta}_{ij} := \frac{\hat{\alpha}_{ij}^2}{\hat{\sigma}_i^2/\lambda_j}, \quad i = 1, \dots, N, \quad j = 1, \dots, p$$

(Step3) 相対効率の推定とリッジ係数の選択:

$$\widehat{\mathcal{R}}_i(k) := \sum_{j=1}^p w_j \frac{\lambda_j^2}{(\lambda_j + k)^2} + \sum_{j=1}^p w_j \widehat{\delta}_{ij} \frac{k^2}{(\lambda_j + k)^2} \longrightarrow \min_{k>0} \implies \widehat{k}_i,$$

$$i = 1, \dots, N$$

(Step4) ORR 推定量の計算:

$$\widehat{\alpha}_j(\widehat{k}_i) := \frac{\lambda_j}{\lambda_j + \widehat{k}_i} \widehat{\alpha}_{ij}, \quad i = 1, \dots, N, \quad j = 1, \dots, p$$

5.3 評価基準

シミュレーション結果の評価基準としては以下の経験総平均 2 乗誤差 (Empirical Total Mean Square Error: ETMSE)¹³⁾ と経験相対効率 (Empirical Relative Efficiency: EREff) を利用した:

OLS 推定量:

$$\text{ETMSE}(\widehat{\alpha}_p) := \frac{1}{N} \sum_{i=1}^N \|\widehat{\alpha}_{p(i)} - \alpha_p\|^2,$$

$$\text{EREff}(\widehat{\alpha}_p | \widehat{\alpha}_p) := \frac{\text{ETMSE}(\widehat{\alpha}_p)}{\text{ETMSE}(\widehat{\alpha}_p)} (= 1)$$

ORR 推定量:

$$\text{ETMSE}(\widehat{\alpha}_p(\widehat{k})) := \frac{1}{N} \sum_{i=1}^N \|\widehat{\alpha}_p(\widehat{k}_i) - \alpha_p\|^2,$$

$$\text{EREff}(\widehat{\alpha}_p(\widehat{k}) | \widehat{\alpha}_p) := \frac{\text{ETMSE}(\widehat{\alpha}_p(\widehat{k}))}{\text{ETMSE}(\widehat{\alpha}_p)}$$

ここで、 $k = \overline{k}, \widehat{k}, \underline{k}$ に関しても同様に定義される。また、 $\|\mathbf{x}\| := \sqrt{x_1^2 + \dots + x_n^2}$ はベクトル $\mathbf{x} = [x_1, \dots, x_n]'$ のノルムであり、

$$\widehat{\alpha}_{p(i)} := \begin{bmatrix} \widehat{\alpha}_{i1} \\ \vdots \\ \widehat{\alpha}_{ip} \end{bmatrix}, \quad \widehat{\alpha}_p(\widehat{k}_i) := \begin{bmatrix} \widehat{\alpha}_1(\widehat{k}_i) \\ \vdots \\ \widehat{\alpha}_p(\widehat{k}_i) \end{bmatrix}$$

とおいた。

13) 総算術平均 2 乗誤差 (Total Average Square Error: TASE) と呼ばれる場合もある。

5.4 結果と考察

各データセットに対する OLS 推定量と ORR 推定量の TMSE と相対効率 (REff) の真値を表 3, 4 に与えており, モンテカルロ・シミュレーションの実行によって得られた ETMSE と経験相対効率 (EREff) の値を表 5, 6 に与える.

表 3: データセット (DS1) に対する OLS 推定量と ORR 推定量の TMSE と相対効率の真値

Estimators	TMSE	REff
OLS	34.518	1.000
ORR(k^*)	6.415	0.186

表 4: データセット (DS2) に対する OLS 推定量と ORR 推定量の TMSE と相対効率の真値

Estimators	TMSE	REff
OLS	83.688	1.000
ORR(k^*)	7.079	0.085

表 5: データセット (DS1) に対するシミュレーション結果

Estimators	ETMSE	EREff
OLS	34.527	1.000
ORR (\hat{k})	18.488	0.535
ORR (\bar{k})	14.279	0.414
ORR ($\hat{\hat{k}}$)	12.185	0.353
ORR ($\bar{\bar{k}}$)	9.171	0.266
ORR(DSW)	7.769	0.225

表 6: データセット (DS2) に対するシミュレーション結果

Estimators	ETMSE	EREff
OLS	83.866	1.000
ORR (\hat{k})	44.483	0.530
ORR (\bar{k})	29.790	0.355
ORR ($\hat{\hat{k}}$)	23.617	0.282
ORR ($\bar{\bar{k}}$)	14.819	0.177
ORR(DSW)	9.266	0.110

これらの結果から以下のことがわかる:

(R1) OLS 推定量の評価基準に関する真値とシミュレーションの結果の表を比べることによって, ある程度の精度が確保されていることがわかる. たとえば, (DS1) に関しては表 3 と表 5 を比較することによって, 小数点第 1 位までの精度が保障されていることがわかる.

(R2) 真値に関する表 3, 4 から, リッジ係数 k を適切に選択することができれば, OLS 推定量を大幅に改良する ORR 推定量を構成できることが示唆される. とくに, (DS2) の場合には, OLS 推定量に対して ORR 推定量が理論的には約 9% まで改良することができる余地があることに注意しよう.

(R3) OLS 推定量と ORR 推定量のシミュレーション結果を比較することによって、全ての場合に関して ORR 推定量の方がよい結果を与えている。特に、Dempster ら (1977) のアルゴリズム (DSW) がもっともよい結果を与えており、次いで $k = \bar{k}$ に関するアルゴリズムがよい結果を与えている。ただし、注意 4.5 でも述べたけれども、Dempster ら (1977) によるアルゴリズムは k に関する候補点の取り方に強く依存することに注意が必要である¹⁴⁾。

(R4) データセット (DS1), (DS2) に対するシミュレーションにおいて選択されたリッジ係数 k のボックスプロット (図 6, 7) からわかるように、TMSE を最小にするリッジ係数の値 k^* (図中では垂直線で表されている。) を偏りなく「当てている」アルゴリズムがない。すなわち、厳密な意味でリッジ係数の理想的な値を再現するアルゴリズムは今回考察したものの中には存在しない。

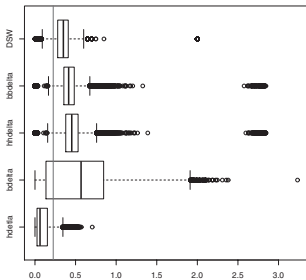


図 6: (DS1) に対するシミュレーションにおいて選択された k のボックスプロット: (垂直線 $k = k^* = 0.225$)

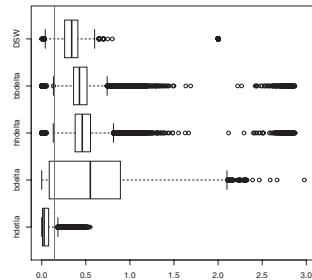


図 7: (DS2) に対するシミュレーションにおいて選択された k のボックスプロット: (垂直線 $k = k^* = 0.145$)

14) ここで与えられたシミュレーションで採用された Dempster ら (1977) によるアルゴリズム用の候補点 (メッシュ) は以下のようなものである:

$k = 0, 0.001, 0.001, 0.002, (0.002), 0.02, 0.03, (0.01), 0.1, 0.11, (0.01), 0.5, 0.6, (0.05), 1.0, 1.1, (0.1), 2.0$

なお, (...) はその左右の数値間の刻み幅を表す。

注意 5.1 図 6, 7 におけるボックスプロットは, シミュレーション回数 $N = 10^6$ の中からランダムに 10^5 個のものを抽出したものを利用して描いていることに注意しよう¹⁵⁾. また, プロットに関する種類を表すラベルと δ_i の推定量, または k の選択アルゴリズムと表 7 のように対応することに注意しよう. なお, DSW は Dempster ら (1977) による k のアルゴリズムによって選択されたものを表す.

表 7: 図 6, 7 におけるラベルと非心度, リッジ係数の対応

ラベル	δ_i	k
hdelta	$\widehat{\delta}_i$	\widehat{k}
bdelta	$\overline{\delta}_i$	\overline{k}
hhdelta	$\widehat{\widehat{\delta}}_i$	$\widehat{\widehat{k}}$
bbdelta	$\overline{\overline{\delta}}_i$	$\overline{\overline{k}}$

6 おわりに

本稿では, 相対効率にもとづくリッジ係数の選択アルゴリズムについて議論してきた. ここで検討されたアルゴリズムのうち, 事前にリッジ係数の候補点を選ばなければならない Dempster ら (1977) のアルゴリズムを除くと, 非心度の推定に関して OLS 推定量と ORR 推定量を折衷したものがもっとも良いという結果を得た. ただし, 厳密な意味でリッジ係数の理想的な値を再現するアルゴリズムは今回考察したものの中には存在しないため, 更なる改良の余地があることが今後の課題として認識された.

また, 井上 (1983) (p. 53) にも指摘されているように, これらのアルゴリズムによって選択されるリッジ係数は標本に依存しており, それ自体が確率変数となるため, 実際に選択されたりッジ係数をプラグインした ORR 推定量, いわゆる実行可能型通常リッジ回帰 (Feasible Ordinary Ridge Regression: FORR) 推定量の解析的に正確なモーメントとそれらから導かれる TMSE や相対効率

15) ファイルサイズの関係上, このような措置と行ったけれども, 分布の本質は失われないように配慮されていることに注意しよう.

などには言及していない。この問題についても今後の課題としたいが、Dwivedi ら (1980) や Jimichi (2005, 2008) において与えられた結果¹⁶⁾ が、この問題を解決するための手がかりになるものと思われる。

謝辞

本研究の一部は文部科学省科学研究費 (基盤研究 (C) 課題番号: 22540162, 研究代表者: 北原和明) の助成により行った。ここに感謝の意を表するものである。

参考文献

- [1] Dempster, A. P., M. Schatzoff, and N. Wermuth (1977) *A simulation study of alternatives to ordinary least squares*, Journal of the American Statistical Association, Vol. 72, pp. 77–106.
- [2] Dwivedi, T. D., V. K. Srivastava, and R. L. Hall (1980) *Finite sample properties of ridge estimators*, Technometrics, Vol. 22, pp. 205–212.
- [3] Gibbons, D. G. (1981) *A simulation study of regression estimators*, Journal of the American Statistical Association, Vol. 76, pp. 131–139.
- [4] Hoerl, A. E. (1962) *Application of ridge analysis to regression problems*, Chemical Engineering Progress, Vol. 58, pp. 54–59.
- [5] Hoerl, A. E., and R. W. Kennard (1970) *Ridge regression: biased estimation for nonorthogonal problems*, Technometrics, Vol. 12, pp. 55–67.
- [6] Hoerl, A. E., R. W. Kennard, and K. F. Baldwin (1975) *Ridge regression: some simulations*, Communications in Statistics, Vol. 4, pp. 105–123.
- [7] 井上 勝雄 (1983) 『計量経済学の理論と応用—多重共線性とモデルの計測—』, 有斐閣.
- [8] 地道 正行 (1994) 『Ridge 定数の決定ルール』, Research Reports On Statistics, Osaka University, No. 38.

16) ある種の実行可能型一般化リッジ回帰推定量の正確なモーメントにもとづく平均 2 乗誤差の評価に関するもの。

- [9] 地道 正行 (2002) 『パラメトリックブートストラップ法を用いた一般化 Ridge 回帰推定量の誤差評価』, 商学論究, 第 50 巻第 1, 2 号 合併号, 関西学院大学商学研究会, pp. 511–526.
- [10] 地道 正行 (2004) 『パラメトリックブートストラップ法を用いた通常リッジ回帰推定量と一般化リッジ型推定量の総平均 2 誤差推定』, 商学論究, 第 52 巻第 2 号, 関西学院大学商学研究会, pp. 41–66.
- [11] Jimichi, M. (2005) *Improvement of regression estimators by shrinkage under multicollinearity and its feasibility*, Ph.D. dissertation, Osaka University.
- [12] Jimichi, M. (2008) *Exact moments of feasible generalized ridge regression estimator and numerical evaluations*, Journal of the Japanese Society of Computational Statistics, Vol. 21, pp. 1–20.
- [13] Jimichi, M., and N. Inagaki (1993) *Centering and scaling in ridge regression*, Statistical Sciences and Data Analysis, Editors K. Matusita, M. Puri, and T. Hayakawa, VSP International Science Publishers, pp. 77–86.
- [14] Johnson, N. L., S. Kotz and N. Balakrishnan (1995) *Continuous Univariate Distributions*, Vol. 2, Second Edition, John Wiley & Sons, Inc.
- [15] Lee, T. S., and D. B. Campbell (1985) *Selecting the optimum k in ridge regression*, Communications in Statistics, Theory and Method, Vol. A14, pp. 1589–1604.
- [16] Nordberg, L. (1982) *A procedure for determination of a good ridge parameter in linear regression*, Communications in Statistics, Simulation and Computation, Vol. 11, pp. 285–309.
- [17] Perlman, M. D., and U. A. Rasmussen (1975) *Some remarks on estimating a noncentrality parameter*, Communications in Statistics, Vol. 4, pp. 435–468.

付録

A 非心カイ自乗分布と非心エフ分布

A.1 非心カイ自乗分布

確率変数 X_1, \dots, X_n がそれぞれ独立に正規分布 $N(\mu_i, 1)$ に従うとき、

$$X := X_1^2 + \dots + X_n^2$$

で定義される確率変数 X の分布は自由度 n , 非心度 (non-centrality parameter) $\delta := \sum_{i=1}^n \mu_i^2$ の非心カイ自乗分布 (non-central chi square distribution) と呼ばれ、

$$X \sim \chi_n^2(\delta) = N(\mu_1, 1)^2 + \dots + N(\mu_n, 1)^2$$

と記号的に表わされる。

自由度 n , 非心度 δ の非心カイ自乗分布の確率密度関数は、

$$f(x; n, \delta) = \sum_{j=0}^{\infty} p(j; \delta/2) f(x; n+2j), \quad 0 < x < \infty, n \in \mathbb{N}, \delta > 0 \quad (18)$$

で与えられる。ここで

$$f(x; n+2j) := \frac{1}{2^{\frac{n+2j}{2}} \Gamma\left(\frac{n+2j}{2}\right)} x^{\frac{n+2j}{2}-1} e^{-\frac{x}{2}};$$

自由度 $n+2j$ のカイ自乗分布 χ_{n+2j}^2 の確率密度関数

$$p(j; \delta/2) := e^{-\delta/2} \frac{(\delta/2)^j}{j!}; \quad \text{ポアソン分布 Po}(\delta/2) \text{ の確率関数}$$

である。

非心カイ自乗分布の平均、分散は以下のように与えられる:

$$E(X) = n + \delta, \quad V(X) = 2n + 4\delta$$

A.2 非心エフ分布

確率変数 X が自由度 m , 非心度 δ の非心カイ自乗分布 $\chi_m^2(\delta)$, 確率変数 Y が自由度 n のカイ自乗分布 χ_n^2 にそれぞれ独立に従うとき、

地道：相対効率にもとづくリッジ係数の選択アルゴリズム

$$F := \frac{X/m}{Y/n}$$

で定義される確率変数 F の分布は、自由度 (m, n) 、非心度 δ の非心エフ分布 (non-central F distribution) と呼ばれ、

$$F \sim F_n^m(\delta) = \frac{\chi_m^2(\delta)/m}{\chi_n^2/n}$$

と記号的に表わされる。

自由度 (m, n) 、非心度 δ の非心エフ分布 $F_n^m(\delta)$ の確率密度関数は、

$$f(x; m, n, \delta) := f(x; m, n) \sum_{j=0}^{\infty} p(j; \delta/2) \left(\frac{mx}{n+mx} \right)^j \frac{B\left(\frac{m}{2}, \frac{n}{2}\right)}{B\left(\frac{m}{2} + j, \frac{n}{2}\right)}$$

で与えられる。ここで

$$f(x; m, n) = \frac{1}{B\left(\frac{m}{2}, \frac{n}{2}\right)} \left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1} \left(1 + \frac{m}{n}x\right)^{-\frac{m+n}{2}}$$

；自由度 (m, n) のエフ自乗分布 F_n^m の確率密度関数

$$p(j; \delta/2) = e^{-\frac{\delta}{2}} \frac{(\delta/2)^j}{j!}; \text{ポアソン分布 } \text{Po}(\delta/2) \text{ の確率関数}$$

である。

非心エフ分布の平均、分散は以下のように与えられる：

$$E(F) = \frac{n}{n-2} \left(1 + \frac{\delta}{m}\right), \quad (n \geq 3), \quad (19)$$

$$V(F) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)} \left\{ \frac{\delta^2}{m(m+n-2)} + \frac{2}{m}\delta + 1 \right\}, \quad (n \geq 5) \quad (20)$$

B 非心度の推定に ORR 推定量を利用した場合の相対効率の推定量の導関数

ORR 推定量 $\hat{\alpha}_i(k)$ を利用した相対効率の推定量

$$\widehat{\mathcal{R}}(k) = \sum_{i=1}^n \frac{w_i(\lambda_i + \widehat{\delta}_i k^2)}{(\lambda_i + k)^2}$$

の 1, 2 次導関数は、以下のように与えられる：

$$\begin{aligned} \frac{d}{dk} \widehat{\mathcal{R}}(k) &= -2 \sum_{i=1}^n \frac{w_i}{(\lambda_i + k)^3} \left(\widehat{\delta}_i k^2 - \lambda_i \widehat{\delta}_i k + \lambda_i^2 \right) \\ &= -2 \sum_{i=1}^n \frac{w_i \lambda_i^2}{(\lambda_i + k)^5} \left(\widehat{\delta}_i k^2 - \lambda_i \widehat{\delta}_i k + (\lambda_i + k)^2 \right) \\ \frac{d^2}{dk^2} \widehat{\mathcal{R}}(k) &= 2 \sum_{i=1}^n \frac{w_i}{(\lambda_i + k)^4} \left(3\widehat{\delta}_i k^2 - 6\lambda_i \widehat{\delta}_i k + \lambda_i^2 \widehat{\delta}_i + 3\lambda_i^2 \right) \\ &= 2 \sum_{i=1}^n \frac{w_i \lambda_i^2}{(\lambda_i + k)^6} \left(3\widehat{\delta}_i k^2 - 6\lambda_i \widehat{\delta}_i k + \lambda_i^2 \widehat{\delta}_i + 3(\lambda_i + k)^2 \right) \end{aligned}$$

ここで,

$$\widehat{\alpha}_i(k) = \frac{\lambda_i}{\lambda_i + k} \widehat{\alpha}_i, \quad \widehat{\delta}_i = \frac{\widehat{\alpha}_i^2(k)}{\widehat{\sigma}^2/\lambda_i} = \frac{\lambda_i^2}{(\lambda_i + k)^2} \widehat{\delta}_i$$

より,

$$\frac{d\widehat{\delta}_i}{dk} = -\frac{2}{\lambda_i + k} \widehat{\delta}_i = -\frac{2\lambda_i^2}{(\lambda_i + k)^3} \widehat{\delta}_i$$

が成り立つことを使った。なお、相対効率の推定量

$$\overline{\mathcal{R}}(k) = \sum_{i=1}^n \frac{w_i (\lambda_i + \overline{\delta}_i k^2)}{(\lambda_i + k)^2}$$

の導関数についても同様に与えられる。

C 計算機環境

計算機環境に関する主な仕様は以下の通りである:

Item	Specification
Central Processing Unit (CPU)	Intel® Core™ i7-3930K (3.20GHz/6 cores)
Memory	DDR3-SDRAM 16GB
Operating System (OS)	Microsoft Windows® 7 Professional (64 bit version)
Software	R for Windows 2.4.1 (64 bit version)
Physical Random Number Generator	Nippon Techno Lab Inc. Random Streamer RPG102